Les métadonnées archivistiques en transition vers des graphes de données : 2 – point d'actualité et précisions

par Florence Clavaud (Archives nationales)

Article publié sur le carnet de recherche des Archives nationales en juillet 2021 (https://labarchiv.hypotheses.org/1728).

Près d'un an et demi après notre premier billet dans ce carnet sur le thème de la transition des métadonnées archivistiques vers des graphes d'entités liées (https://labarchiv.hypotheses.org/1495), il nous a paru nécessaire de faire un **point d'actualité sur *Records in Contexts* (RiC)**.

Sur cette nouvelle norme du CIA qui à ce stade reste peu connue, et sur le contexte technologique dans lequel elle s'inscrit, beaucoup de choses ont été dites ou écrites depuis février 2020, notamment en France, qui ont pu soulever diverses interrogations voire créer de la confusion. Notre objectif sera donc ici aussi **d'apporter quelques éléments de clarification afin de nourrir les réflexions de chacun et de favoriser des discussions constructives**.

Nous parlons ici à la fois en tant que membre exécutif, depuis sa création en 2012, du groupe de travail EGAD du CIA qui est l'auteur de RiC, en tant que responsable de l'équipe de développement de l'ontologie RiC-O au sein de ce groupe, et en tant que responsable de la Mission référentiels des Archives nationales (AN), avec une expérience multiforme de plus de 25 ans en terme de modélisation, de production, de gestion de la qualité, de publication et de retraitement des métadonnées archivistiques, ou de conduite de projets numériques innovants.

Merci beaucoup à Pauline Charbonnier, qui nous a aidée à finaliser cet article et surtout qui travaille avec nous depuis plus de trois ans maintenant.

Sommaire :

1. Point d'actualité : où en est RiC ? 2
2. Qu'est-ce que RiC ? 3
3. Qu'est-ce que RiC change ? 7
4. Comment réaliser la transition vers RiC, notamment à partir d'instruments de recherche en XML/EAD ? 12
5. Est-ce que RiC reste pure théorie aujourd'hui, ou bien y a-t-il déjà des réalisations concrètes, utilisant RiC-O et RDF ? 14
6. Mais finalement à quoi RDF/RiC-O peut-il très concrètement être utile pour un service d'archives ? 19
7. En conclusion (de cet article) 23

1. Point d'actualité : où en est RiC ?

Depuis septembre 2016 (date de la publication de RiC-CM 0.1) et décembre 2019 (publication de RiC-O 0.1) les travaux du groupe EGAD ont bien avancé.

Depuis mars 2020, l'ontologie RiC-O est dotée d'un entrepôt de sources public sur GitHub (https://github.com/ICA-EGAD/RiC-O). Dans cet entrepôt, on trouve le fichier source de RiC-O dans ses différentes versions successives ainsi que des exemples et diagrammes ; **on peut aussi y échanger en permanence avec l'équipe de développement de RiC-O et y soumettre des tickets** ('issues' : questions, signalements de problèmes ou d'anomalies, propositions d'évolution), ou y faire des commentaires sur les tickets ou travaux en cours. RiC-O dispose également d'un site web d'information (https://ica-egad.github.io/RiC-O/), sans compter une liste de discussion. En outre, l**a version 0.2 de RiC-O a été publiée en février 2021** (https://www.ica.org/standards/RiC/ontology).

RiC-O 0.2 est conforme à la version 0.2 du modèle conceptuel RiC-CM que le groupe EGAD a publiée le 22 juillet 2021 (voir https://www.ica.org/fr/publication-de-records-in-contexts-conceptual-model-02). RiC-CM 0.2 présente quelques différences notables avec sa prévisualisation publiée en décembre 2019 (présence d'une introduction de taille significative ; changements ponctuels dans la définition de certaines entités ; changement du nom d'un nombre important de relations pour prendre en compte les relations passées).

Un appel à commentaires sur RiC-CM 0.2 a également été lancé le 22 juillet 2021 (voir https://www.ica.org/fr/appel-a-commentaires-sur-ica-ric-cm-02) . Les très nombreux commentaires reçus de septembre 2016 à janvier 2017 sur RiC-CM 0.1 ont été extrêmement utiles à EGAD pour l'élaboration de RiC-CM 0.2. Le groupe EGAD espère que cette nouvelle étape d'appel à commentaires sera tout aussi fructueuse et utile. Ce nouveau standard international de description des archives ne peut être élaboré qu'**avec et par** les communautés professionnelles (archivistes, record managers, ingénieurs, réutilisateurs des métadonnées) auxquelles il s'adresse. **L'enjeu est important puisqu'il s'agit de finaliser, pour la publier en 2022, une version 1.0 de RiC-CM et de RiC-O, qui n'aura plus le statut de projet (« draft ») mais de recommandation normative remplaçant les quatre normes actuellement en vigueur pour la description des archives (ISAD(G), ISAAR(CPF), ISDF et ISDIAH.**

Autrement dit, ce **n'est pas dans un futur lointain que les archivistes auront à s'intéresser à RiC, c'est dès maintenant ou au plus tard en 2022.**

Pour faciliter la compréhension et l'appropriation de RiC en France, nous organisons notamment une série de trois webinaires en septembre et octobre 2021, en collaboration avec le SIAF :

- Le jeudi 30 septembre de 10 h à 11 h 30 : Présentation générale de RiC et de ses enjeux (lien d'inscription : https://www.eventbrite.fr/e/1648,44178375)
- Le jeudi 7 octobre de 10 h à 11 h 30 : Les entités dans RiC (lien d'inscription : https://www.eventbrite.fr/e/164855135147)
- Le vendredi 15 octobre de 10 h à 11 h 30 : Les relations dans RiC (lien d'inscription : https://www.eventbrite.fr/e/164856264525).

Un programme plus précis sera diffusé prochainement. Il est déjà possible de s'inscrire aux séances (inscriptions possibles pour une ou plusieurs séances).

2. Qu'est-ce que RiC ?

Dans sa version 0.2, RiC est constitué d'un modèle conceptuel abstrait de type entités-relations (RiC-CM) et d'une ontologie (RiC-O), écrite dans les langages OWL, RDFS et RDF, transposition technique de RiC-CM directement utilisable pour produire des jeux de métadonnées archivistiques en RDF. Bientôt, à ces deux parties viendra s'ajouter une introduction générale (RiC-IAD). Enfin, à une date qu'il est difficile de préciser pour l'instant, le groupe EGAD ajoutera à ces trois composants un guide de mise en œuvre (RiC-AG), qui sera probablement composé de divers documents.

Comme les quatre normes précédentes, RiC s'applique à n'importe quel type de ressources archivistiques, analogiques, nativement numériques ou hybrides, quels que soient leur provenance, leur nature et leur date.

RiC est comme les précédentes normes du CIA une norme générique et internationale ; on n'y trouvera donc pas toujours directement, clé en main, de quoi satisfaire des besoins de description très spécifiques, liés à des documents ou à un contexte particulier.

Le périmètre de RiC, ce sont les métadonnées de description des archives. Le périmètre ne change donc pas par rapport aux précédentes normes. Mais il est tout de même aujourd'hui utile de préciser qu'il ne faut pas confondre les métadonnées de description des archives avec les quantités massives

de *données* nativement numériques, hétérogènes, aux structures et formats sans cesse en évolution, qui entrent dans les SAE, ni avec les métadonnées très diverses qui accompagnent les items constituant les données ou documents nativement numériques, ou qui sont embarquées dans ces archives.

Il ne s'agit pas non plus des métadonnées qui servent à la gestion matérielle et à la préservation des archives. Cependant des points de chevauchement conséquents existent entre ces catégories de métadonnées : par exemple, l'état matériel ou la volumétrie des documents, le format des fichiers numériques, les dimensions des documents analogiques, les règles de communication, de diffusion ou de réutilisation applicables, sont aussi considérées depuis longtemps comme des métadonnées descriptives, et RiC les inclut dans son périmètre.

L'objet de RiC, ce sont les données que les services d'archives ont besoin de consolider ou de produire, automatiquement ou semi-automatiquement (et dans ce cas ce sera à partir des archives nativement numériques, ou des documents numérisés) ou non, puis de gérer, au sujet de l'ensemble des archives qu'ils conservent, pour les identifier, les signaler et les rendre accessibles aux publics (nous en reparlons plus loin). Ces métadonnées de description sont consubstantielles aux services d'archives ; et elles ont besoin d'un modèle pertinent, expressif et partagé.

Voilà pour les points de ressemblance.

Mais RiC se distingue des précédentes normes, d'abord parce qu'il s'agit d'un **cadre de référence global et unique**. Son ambition est de **faciliter, pour les professionnels du patrimoine culturel notamment, la compréhension du domaine métier dans son ensemble, et la mise en place d'infrastructures de qualité pour la description des archives. Si RiC est appliqué par un nombre de plus en plus grand d'institutions et de projets, il facilitera le partage de savoir-faire et d'outils, l'interconnexion et l'échange de métadonnées, la construction d'applications et de portails communs.** C'est ce à quoi servent les normes professionnelles de description dans le monde des institutions culturelles ; elles font gagner du temps dans les étapes de modélisation et de conception des systèmes d'information et des logiciels, elles aident à définir les stratégies et dispositifs locaux et facilitent l'interopérabilité entre les applications. **Pour les utilisateurs finaux, quels qu'ils soient, les gains devraient également être importants – nous évoquons ce sujet plus loin (dans la partie Qu'est-ce que RiC change ?).**

Par ailleurs, dans RiC il est principalement question, non pas de principes et de règles applicables à la description des archives, mais des catégories d'objets (des types d'entités ; il y a 22 entités dans RiC-CM 0.2) que l'on peut distinguer dans l'univers des archives, des caractéristiques de ces entités (leurs attributs, au nombre de 41 dans RiC-CM 0.2) et des relations que ces entités peuvent avoir entre elles (au nombre de 78 dans RiC-CM 0.2, sans compter les relations inverses). Soit bien plus de composants que les précédentes normes ne le spécifiaient de façon par ailleurs plus ou moins vague ou implicite.

En effet, RiC **considère les métadonnées archivistiques comme un graphe d'entités liées, de même que le modèle conceptuel IFLA LRM (Library Reference Model, voir https://www.ifla.org/publications/node/11412), publié en 2017, le fait en ce qui concerne les métadonnées bibliographiques**. Il invite donc les archivistes à s'engager dans un processus de transformation des métadonnées de description comparable à celui qui commence dans les bibliothèques – comme en France, où ces opérations sont accompagnées par le programme Transition bibliographique (https://www.transition-bibliographique.fr/) lancé en 2015 pour faciliter et accompagner la mise en œuvre concrète, dans les bibliothèques, de LRM et du nouveau code de catalogage RDA-FR, transposition française de RDA.

Pour les métadonnées archivistiques, il ne s'agit pas seulement de s'inscrire dans une tendance générale fondée sur l'idée que, dans l'état actuel des technologies, les graphes orientés et étiquetés sont la méthode de représentation de données la plus expressive, celle aussi qui démultiplie les possibilités d'exploitation. Il s'agit aussi, beaucoup plus fondamentalement, de traduire pleinement l'affirmation suivante que les archivistes n'ont cessé d'énoncer de diverses manières depuis près de 150 ans : ***les archives*** **ne peuvent pas être comprises (évaluées, prises en charge, classées, décrites, exploitées) sans prendre en compte et comprendre *leurs contextes* de production, de transmission et de gestion**. Cette idée fondamentale est celle qui a conduit à énoncer dans ISAD(G) le principe de la description à plusieurs niveaux et à le décliner en quatre règles. Cependant, dans un instrument de recherche (IR) conforme à ISAD(G), c'est essentiellement le contexte documentaire immédiat qui fait l'objet de toutes les attentions, et de ce fait la bi-dimensionnalité y prédomine. Les niveaux de description correspondent à des **relations d'inclusion ou hiérarchiques**, représentées par des indentations dans un IR au format PDF ou HTML, ou par une arborescence d'éléments XML <c> dans un fichier EAD ; l'ordre de rédaction des descriptions des unités documentaires de même niveau – séries dans un fonds, dossiers dans une série, pièces dans un dossier….- correspond à des **relations séquentielles entre ces unités de description**, déterminées par l'auteur de l'IR**. Dans le modèle RiC, les archives se trouvent explicitement inscrites dans un univers multi-dimensionnel. Les relations hiérarchiques et séquentielles qui existent entre les unités de description y sont clairement spécifiées : elles ne sont plus représentées au travers de conventions de mise en forme ou en s'appuyant sur l'arborescence interne des fichiers**. Et elles figurent parmi bien d'autres. Enfin, toutes ces entités et relations s'inscrivent dans le temps et l'espace. C'est cette idée fondamentale qui a donné son nom à RiC (*Records in Context**s***). **On peut aussi envisager ces couches de contextes interconnectés comme des points de vue différents** : la provenance organique et fonctionnelle, le contexte documentaire (les relations entre documents eux-mêmes), et bien d'autres (par ex. les facettes que la diplomatique distingue, comme l'auteur du document, son destinataire ; celles que l'indexation classique des IR considère, comme les sujets des documents, etc.).

Ainsi, en prenant RiC comme cadre global de référence pour décrire les archives, **on se met en situation de produire des descriptions plus proches d'une réalité complexe et changeante, plus précises, plus riches, plus nuancées, plus exploitables.**

Une autre caractéristique de ce modèle, majeure dans RiC-CM et fondamentale dans RiC-O, est son organisation en poly-hiérarchies. L'entité Agent, par ex., regroupe six sous-entités (Person, Group qui inclut lui-même Family et Corporate Body, Position et Mechanism).

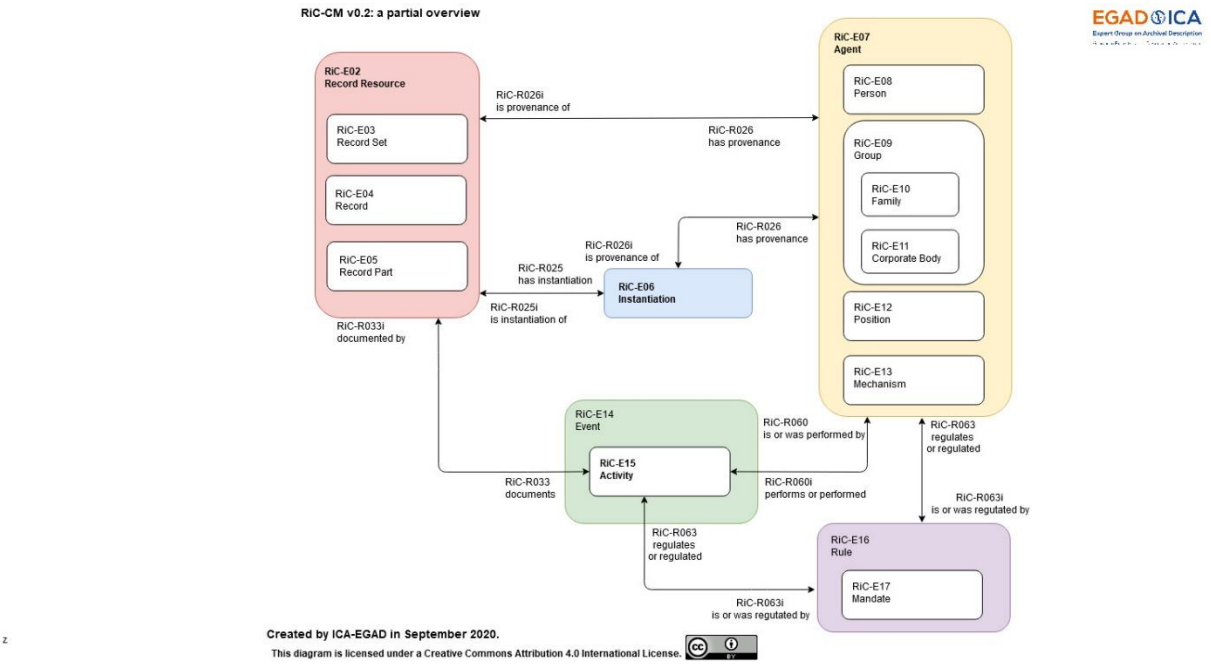


Figure 1 : aperçu de RiC-CM 0.2 (les principales entités et quelques relations)

Cela procure, lorsqu'on implémente RiC, **une grande flexibilité, celle dont on a toujours eu absolument besoin lorsqu'on décrit des archives,** car on le fera plus ou moins précisément en fonction de la nature des ensembles documentaires ou items que l'on décrit, des connaissances que l'on a pu accumuler sur ces documents, des moyens et du temps dont on dispose. Ainsi, lorsqu'on ne sait pas si on a affaire à un Record ou un Record Part ou un Record Set (comme cela peut être le cas à la seule lecture de nombreux inventaires anciens), on peut utiliser la superclasse Record Resource ; inversement l'entité Record a ses propres attributs et est le domaine de certaines relations spécifiques, mais elle hérite aussi des attributs de Record Resource et peut être utilisée avec les relations dont le domaine est Record Resource. Dans RiC-O, on dispose de plus d'options encore, qui permettent d'affiner la description lorsqu'on le souhaite ; on peut ainsi, par ex., utiliser la relation directe '*is or was leader of*', comme celle définie par RiC-CM, pour relier une personne et une collectivité que cette personne dirige ou a dirigé, ou encore choisir d'utiliser une entité Leadership Relation, pour lier la personne et la collectivité via cette entité, ce qui permet d'assigner une date et une description à la relation, comme d'ailleurs on peut le faire en EAC-CPF (voir la figure 2).

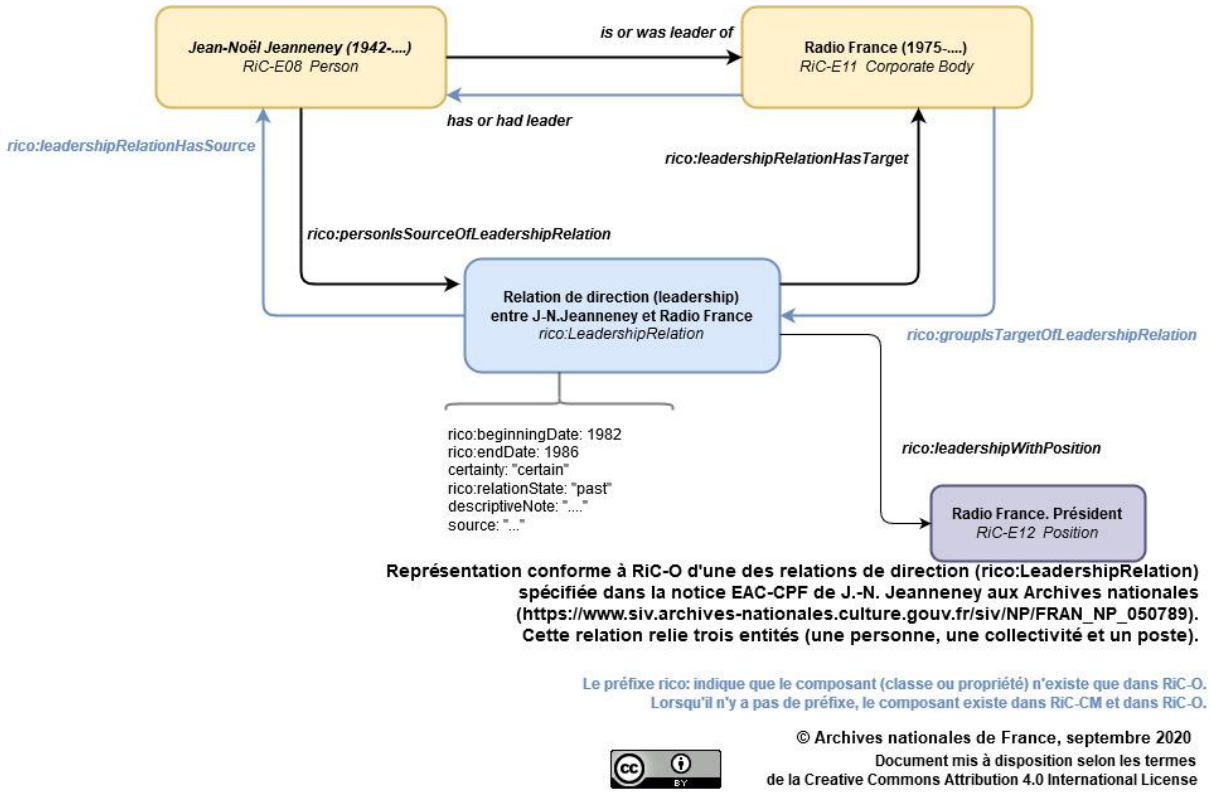


Figure 2 : exemple de représentation d'une relation avec RiC-O 0.2, soit directement avec « is or was leader of » soit en utilisant une instance de la classe LeadershipRelation

Cela facilite aussi le **liage** : les entités et relations de haut niveau, assez génériques, comme *Group*, *Event* ou '*is or was part of*', sont de bons candidats pour des mappings avec des entités ou relations d'autres modèles comme PREMIS, PROV-O, IFLA LRM ou CIDOC CRM, ce qui sera d'ailleurs fait dans les versions ultérieures de RiC-CM et RiC-O.

Enfin, cette organisation poly-hiérarchique facilite les extensions du modèle pour répondre à des besoins spécifiques. Certains attributs, d'ailleurs, sont définis comme extensibles. Si on le souhaite, on peut définir une relation qui n'existe pas dans RiC, mais qui serait nécessaire localement, comme une sous-relation d'une relation RiC ; elle héritera alors de ses caractéristiques, domaine, portée, et pourra toujours être remplacée par la super-relation définie dans RiC.

En d'autres termes, si on choisit de mettre en œuvre RiC, on peut choisir de dépasser le niveau de précision des instruments de recherche ou notices d'autorité actuels ou non, de faire évoluer ses métadonnées voire son modèle local de métadonnées au fil des années et des besoins ou moyens. Ce qui s'avère crucial dans un contexte où les moyens humains et financiers sont contraints.

3. Qu'est-ce que RiC change ?

D'un certain point de vue, cela ne change rien, car RiC se situe dans le prolongement des normes qui l'ont précédé, et ne remet pas en cause les fondements de la description archivistique, que des décennies de pratique ont éprouvés. RiC se positionne, comme indiqué ci-dessus, comme un cadre de référence global qui englobe, articule et précise ce qui existe déjà. Ainsi que le groupe EGAD l'a écrit

dans l'introduction à RiC-CM, **on peut continuer à utiliser ISAD(G) et ISAAR(CPF) ; globalement ce qu'on produira sera conforme à RiC - sans l'utiliser à plein ni donc en tirer tout le potentiel.**

Mais d'un autre côté, cela change tout, ou pour être plus précis cela devrait progressivement tout changer.

Dans la plupart de nos systèmes d'information institutionnels, les métadonnées décrivant les archives sont dispersées et enfermées au sein de silos mis en place au fil de plusieurs décennies d'informatisation et désormais juxtaposés les uns aux autres sans qu'aucun lien ne soit le plus souvent établi entre elles.

La même entité de contexte peut ainsi y être décrite plusieurs fois sans lien entre les descriptions. Ainsi, par exemple, aux AN, l'architecte Henri Labrouste (1801-1875) a été l'objet, dans le référentiel des producteurs, d'une notice d'autorité riche conforme à ISAAR(CPF)[1] *puisqu'il a produit des archives que l'institution conserve et que le référentiel des producteurs a été mis en place pour décrire la provenance organique des fonds*. On trouve d'autres informations sur H. Labrouste au sein de la notice du dossier LH//1420/91[2], dans *la base de données Léonore, instrument de recherche mise en place pour lister et rendre accessibles les dossiers de Légion d'honneur ouverts à l'occasion des promotions dans cet ordre et conservés aux AN*. Ajoutons que, *dans le référentiel utilisable aux AN pour indexer les personnes physiques mentionnées dans les instruments de recherche lorsque celles-ci ne sont pas les producteurs des documents décrits, donc pour satisfaire un autre besoin précis*, il existe une notice sommaire sur Henri Labrouste[3]**.** Le nom « Henri Labrouste » ou une autre forme du nom de l'architecte, associé à diverses informations, est par ailleurs simplement mentionné ou balisé librement, au sein de plusieurs répertoires numériques, dans l'intitulé ou la présentation du contenu

[1] Voir dans la salle des inventaires virtuelle cette notice de producteur https://www.siv.archives-nationales.culture.gouv.fr/siv/NP/FRAN_NP_051483 (que l'on pourra trouver via le formulaire de recherche par producteurs). Cette notice fournit l'identifiant ISNI de l'architecte, ses dates de naissance et de décès, une biographie rédigée, une bibliographie ; elle spécifie des relations entre H. Labrouste et trois collectivités et des relations d'identité avec les notices d'autorité de la BnF et de Wikipédia.

[2] Voir dans Léonore la notice https://www.leonore.archives-nationales.culture.gouv.fr/ui/notice/204228. On y trouve notamment son nom incluant l'ensemble de ses prénoms – Pierre, François et Henri –, son lieu de décès, son identifiant ISNI, l'indication de distinctions reçues, et bien sûr le lien vers le dossier de Légion d'honneur numérisé.

[3] Cette notice très courte, qui n'est pas accessible dans la salle des inventaires virtuelle, contient, de son côté, un lien vers une notice du référentiel VIAF. Elle a été utilisée pour lier le nom de Henri Labrouste, balisé à l'aide de l'élément EAD <persname>, dans la description du dossier coté F/17/7231 dossier 20 (https://www.siv.archives-nationales.culture.gouv.fr/siv/UD/FRAN_IR_050281/c-80fohy0ih-11elu28bvxedt), dans l'inventaire relatif au Personnel du service des édifices diocésains et du service du contrôle des édifices diocésains (1840-1927). Mais elle n'a pas servi dans les autres instruments de recherche mentionnés ci-dessus. Noter que pour diverses raisons, le référentiel des personnes physiques reste peu utilisé et ne contient que 2134 notices, dont de très nombreuses se réduisent à un nom. Mais dans le cadre des travaux que la Mission référentiels mène actuellement pour enrichir et restructurer les référentiels des AN, 102 entités décrites dans ce référentiel ont pu être alignées avec autant d'entités du référentiel des producteurs. Des redondances existent donc bel et bien entre ces deux référentiels, et entre le référentiel des producteurs et celui dit des personnes morales, qui est comparable de par les modalités de son utilisation à celui des personnes physiques. Tant que des référentiels seront construits de façon indépendante et pour remplir des besoins différents, tout en concernant le même type d'entités, redondances et dispersions subsisteront et ne pourront que se multiplier. C'est pourquoi la création d'un référentiel unique des agents a été décidée aux AN.

de plusieurs dossiers[4]. **Mais aucune relation ne vient établir dans le système d'information que toutes ces notices et mentions concernent une seule et même personne**.

On le voit par cet exemple très simple et spécifique mais significatif, l'existence de divers silos induit une dispersion des métadonnées de contexte, et des redondances. Elle augmente les risques d'incohérence ou d'inexactitude en empêchant les auteurs des métadonnées d'accéder facilement à l'ensemble des connaissances dont leur institution dispose sur ces entités. Elle crée au minimum de l'ambiguïté pour les utilisateurs des métadonnées (ainsi par exemple, comment être certain que les répertoires numériques mentionnés ci-dessus, la notice de Léonore et la notice de producteur, évoquent bien la même personne, l'architecte et non un homonyme ? les éléments de contexte, dates, noms, indications de profession, etc. fournis par ces inventaires peuvent certes aider à le prouver *pour un lecteur humain assez averti* ; mais bien évidemment ce ne sera pas toujours le cas.

En fait, ce phénomène n'affecte pas que les entités contextuelles (les lieux, les personnes physiques ou morales, les activités dont résultent les documents, les types de documents, les sujets des documents…), **il peut aussi concerner les documents eux-mêmes**. **Aux AN en particulier, le même document ou le même dossier sont susceptibles d'être aujourd'hui décrits plusieurs fois dans plusieurs entrepôts ou instruments de recherche**, car des générations d'archivistes ont successivement travaillé sur les mêmes fonds et séries entrés aux AN dans les deux siècles passés, en employant des moyens et des méthodes différents et pour répondre à des besoins différents (analyse détaillée d'un fonds ancien, répertoire méthodique, base de données thématique…), dans tous les cas en s'appuyant sur la connaissance qu'ils avaient des documents au moment où ils ont travaillé. Nous allons nous employer à quantifier précisément ce phénomène pour les AN. Quoi qu'il en soit, même si on a pu patiemment créer des liens entre les composants de ces instruments de recherche (via l'élément EAD <otherfindaid>), ou entre instruments de recherche synthétiques et analytiques, du point de vue de l'utilisateur cette *stratification des métadonnées, combinée avec le caractère hiérarchique des instruments de recherche classiques, rend souvent difficile à comprendre une liste de résultats de recherche classique et finit par* ***masquer la réalité documentaire que ces métadonnées sont censées révéler***.

Enfin, il est classique, dans nos systèmes, que chacun des silos de métadonnées, des plus anciens, parfois vieillissants, jusqu'aux plus récents comme les SAE, dispose de sa propre interface de recherche et de consultation, parfois non accessible directement au grand public. Les silos ont été dotés d'interfaces spécifiques au moment de leur création ou un peu après ; il n'y a donc le plus souvent pas d'interface de recherche et de consultation unique. C'est vrai aux AN[5] et ailleurs - ailleurs que dans les services d'archives également. La conséquence, pour revenir à notre exemple, c'est qu'il faut faire plusieurs recherches pour disposer de l'ensemble des résultats en ce qui concerne les documents associés à Henri Labrouste.

Finalement, **pour tous les utilisateurs y compris l'expert rompu à la recherche en archives ou l'archiviste**, cette situation brouille aujourd'hui les cartes. Elle **crée au minimum de la confusion, elle induit le plus souvent des problèmes importants de recherche** (du silence ou du bruit). De façon générale, elle ajoute de la complexité à un univers déjà par essence complexe.

---

[4] Voir dans la SIV des AN, la liste des 20 résultats de recherche obtenus après avoir saisi « Henri Labrouste » dans le formulaire de recherche avancée « recherche dans tous les inventaires » (zone de saisie recherche libre, expression exacte).

[5] La SIV intègre cependant à ses listes de résultats des références aux bases de données externes que sont Léonore, ARCADE et ARCHIM.

En rupture avec cette approche, RiC invite à *regrouper les données décrivant chacune des entités présentes et reconnues comme intéressantes pour les besoins du métier, afin de construire des graphes dans lesquels chacune de ces entités est représentée par un nœud*. On peut alors envisager pour le SI local une architecture orientée données[6], rendant ces graphes de métadonnées accessibles aux diverses applications constituant le SI en fonction des besoins à satisfaire (des différents dispositifs de gestion des entrées, des articles ou des communications, aux outils de production et de publication des métadonnées descriptives en passant par la bibliothèque numérique).

Le premier bénéfice que l'on peut attendre de cette transformation est, au sein du service qui l'aura menée à bien, de gagner considérablement en efficacité dans la gestion des métadonnées, la rationalisation de leur production et la réduction des redondances ; du fait que l'on aura agrégé ou lié ces données, **on pourra bien évidemment aussi consolider voire enrichir les connaissances détenues par le service**.

Une telle transformation demande de **travailler en profondeur car il s'agit de faire émerger du contenu des divers entrepôts de métadonnées les entités qui y sont souvent représentées** (les documents, les groupes de documents, leurs différentes manifestations physiques en incluant les reproductions numériques des documents analogiques, mais aussi de nombreux types d'entités contextuelles en incluant celles qui ne sont encore que peu voire pas décrites dans nos vocabulaires ou référentiels documentaires, des règles de communicabilité[7] aux formats en passant par les catégories de fonctions et les types de groupes de documents[8]) **et les relations qui lient les documents à ces entités**. Une telle entreprise ne peut être menée que **par étapes successives**, après avoir élaboré une stratégie tenant compte de l'existant et de ses principaux défauts (des audits sont nécessaires), définissant un modèle local de métadonnées à partir du cadre général de référence RiC, déterminant ce qui doit être fait en priorité, mettant en place des chaînes de traitement dans lesquelles il faudra combiner des processus automatiques et des étapes de validation humaine. Il faut aussi se préoccuper

---

[6] Comme le prônent les signataires du manifeste « Data Centric Manifesto » (http://www.datacentricmanifesto.org/); comme l'INA, par ex., est en train de le faire (à ce sujet on peut consulter l'enregistrement de la présentation faite en décembre 2020 par Gautier Poupeau dans le cadre du séminaire *Les nouveaux paradigmes de l'archive* : https://nparchive.hypotheses.org/738), ou comme le RijkMuseum le prévoit aussi (cf. https://theartofinformationblog.wordpress.com/2020/11/06/integration-layer/).

[7] Les règles de communicabilité, de diffusion et de réutilisation sont désormais parfois formalisées au sein de référentiels *de gestion*, souvent *uniquement utilisables au sein des SAE*. D'un autre côté, les instruments de recherche archivistiques classiques, conformes à ISAD(G) ou à EAD, spécifient en principe les modalités d'accès, de diffusion et de réutilisation des documents, mais le plus souvent sous la forme de texte libre. RiC considère bien évidemment ces règles comme entrant dans son périmètre et définit notamment une entité Rule, ses attributs et relations. RiC-0 ajoute la classe (le type d'entité) RuleType. On dispose donc d'un modèle qui peut aider à construire localement des référentiels mutualisés pour de telles règles, en les alignant le cas échéant avec des référentiels nationaux comme la liste contrôlée des règles de communicabilité établie et publiée par le SIAF au format RDF/SKOS (http://data.culture.fr/thesaurus/resource/ark:/67717/98b3feb3-744b-4ba3-8381-18ca4a555a4c)).

[8] Comme, par exemple, les types de dossiers, dont il pourrait s'avérer utile de produire, pour l'utilisateur final, une description incluant bien sûr une définition mais aussi par ex. la période pendant laquelle tel type de dossier est produit, des liens vers le ou les producteurs susceptibles de l'avoir produit, vers les types de documents que l'on peut s'attendre à y trouver, vers les règles de communicabilité applicables. En d'autres termes, des informations que l'on trouve souvent, pour les dossiers d'archives publiques contemporaines, dans les tableaux de gestion.

d'une stratégie et de moyens d'assignation d'identifiants uniques et pérennes aux entités ainsi isolées[9].

Elle conduit également inévitablement et rapidement à **s'intéresser aux réservoirs de métadonnées externes** : si l'institution ne dispose pas de description de certaines personnes ou lieux ou types de documents mentionnés dans ses instruments de recherche, pour construire ses propres métadonnées sur ces entités, elle aura besoin de jeux de données produites ailleurs et dignes de confiance, autrement dit de référentiels externes. Et réciproquement : ce qu'elle aura produit pourra intéresser d'autres institutions ou projets. En premier lieu d'ailleurs, cela pourra intéresser l'écosystème administratif dans lequel l'institution se situe ; ainsi aux AN, à la fois les missions dans les ministères et le SIAF.

Pour l'utilisateur final, dans l'interface web du SIA, une fois cette transformation opérée, on peut compter **améliorer la performance des moteurs d'indexation et de recherche** (diminuer le silence car les métadonnées seront plus riches, et le bruit car elles seront plus précisément qualifiées), **construire de nouveaux points d'entrée plus intuitifs que celui des producteurs** (onglets, formulaires et/ou facettes « personnes », « collectivités », « lieux », « fonctions », « types de documents »…), **proposer divers dispositifs d'exploration** (par le biais de listes hiérarchisées pour des concepts comme des types de fonctions, par la visualisation des couches de contexte telles que projections sur une ligne de temps et/ou sur une carte, par l'affichage du réseau formé par les agents en relation avec les archives, ou par les documents en suivant les différentes relations qui les relient, par le passage de la description du contenu intellectuel d'un document à la description de chacune de ses instantiations[10] puis accès à leur instantiation numérique le cas échéant…), **des rebonds vers d'autres systèmes d'information** lorsqu'on aura pu lier la description locale des entités de contexte avec leur description dans d'autres systèmes. Globalement, on peut donc espérer des avancées importantes en termes de clarté, de précision et de fonctionnalités.

Ainsi que les auteurs du modèle IFLA LRM (déjà cité) des métadonnées bibliographiques l'ont écrit[11], **trouver, identifier, sélectionner les documents utiles et y accéder, découvrir (via les relations) des documents devrait être plus facile ; ajoutons à ces cinq verbes 'comprendre'**, puisqu'encore une fois les contextes dans lesquels les archives ont été produites, transmises et se trouvent aujourd'hui sont des clés de compréhension de ces documents.

Toujours pour l'utilisateur final, humain ou machine, **les métadonnées conformes à RiC sont susceptibles de constituer, hors du SI local, des jeux de données à fort potentiel de réutilisation**, puisque chacun des éléments qui les composent sont précisément qualifiés (sémantisés) et que les entités de contexte décrites (en particulier) sont dotées d'identifiants uniques et liables avec d'autres

---

[9] On peut notamment s'appuyer sur le vademecum *Identifiants pérennes pour les ressources culturelles* publié en 2015 par le ministère de la Culture dans le cadre de sa feuille de route stratégique « Métadonnées culturelles et transition web 3.0 »Voir le document de la feuille de route sémantique (https://www.culture.gouv.fr/Sites-thematiques/Innovation-numerique/Donnees-publiques/Identifiants-perennes-pour-les-ressources-numeriques).

[10] RiC-CM définit l'entité Instantiation comme étant l'inscription ou la représentation du message informationnel (que véhicule une ressource archivistique) dans une forme physique (numérique ou analogique). Une ressource archivistique (Record Resource) peut avoir plusieurs instantiations. Par exemple, l'original d'une lettre, si c'est un document analogique, a (ou a eu) une première instantiation sur support papier ; la numérisation de cette instantiation produira au moins une deuxième instantiation (image numérique). Cette définition est proche de celle donnée dans le dictionnaire de données PREMIS à l'entité Representation (de même que la définition de Record Resource est proche de celle de Intellectual Entity dans PREMIS).

[11] IFLA LRM, chapitre 3 (*Users and user tasks*), p. 15-16.

données, issues par ex. de projets de recherche. De tels jeux de métadonnées sont alors réutilisables pleinement pour enrichir ou construire des bases de connaissances produites dans le cadre de projets de recherche ou par des communautés professionnelles ou citoyennes. Rien n'empêchera par la suite l'institution ou le projet qui a produit ces jeux de métadonnées, pourvu qu'elle ait défini un modèle et une architecture assez souples et flexibles pour ce faire – et en tout cas RiC est souple et flexible - , de ramener en son sein tout ou partie des données produites par ces partenaires, qu'elle considérerait utiles pour ses propres besoins.

Les métadonnées archivistiques conformes à RiC auront donc toute chance d'être conformes aux principes FAIR[12].

4. Comment réaliser la transition vers RiC, notamment à partir d'instruments de recherche en XML/EAD ?

Il serait totalement irréaliste d'affirmer que cette transition ne prendra pas de temps. La cible est très différente de ce dont nous disposons aujourd'hui, et la route sera longue. Les professionnels des bibliothèques ont souvent entrepris un peu avant les archivistes le même type de transformation ; à la BnF ce processus prendra plusieurs années[13]. Nous l'avons déjà dit, cela ne peut se faire que par étapes ; mais chacune des étapes conduira à des changements significatifs et les futurs systèmes d'information devraient être très différents.

Faut-il renoncer aux instruments de recherche en XML/EAD ? Ils sont là, c'est un fait. Il est tout aussi impensable de les ignorer que de ne pas réfléchir à l'avenir. Les programmes systématiques, ambitieux et coûteux d'informatisation et de normalisation dont ils sont le résultat sont même une condition indispensable pour la réussite du processus.

Pour préciser un peu les choses, il faut rappeler que **nos métadonnées ont une histoire**. Au tournant des années 2000, pour rester sur le territoire français, les services publics d'archives utilisaient quotidiennement une masse significative d'inventaires papier, imprimés, dactylographiés ou manuscrits (certains microfilmés ou microfichés), de fichiers papier, et des bases de données dites textuelles[14]. Ils avaient hérité cette production scientifique de générations d'archivistes, souvent des

---

[12] *The FAIR Guiding Principles for scientific data management and stewardship* (2016), recommandations pour améliorer "*the Findability, the Accessibility, the Interoperability and the Reuse of digital assets*", autour desquelles une communauté et des initiatives diverses se sont développées. Voir le site web https://www.go-fair.org/.

[13] Voir le site web Transition bibliographique déjà cité (https://www.transition-bibliographique.fr/) et, entre autres présentations, celles faites par Tiphaine-Cécile Foucher et Anila Angjeli lors de la journée d'étude du 28 janvier 2020 (https://labarchiv.hypotheses.org/files/2020/02/20200128_7_TransitionBibliographiqueALaBnF_bd.pdf et https://labarchiv.hypotheses.org/files/2020/02/20200128_8_VersUnFichierNationalDEntites.pdf).

[14] Aux AN cet existant comptait des milliers d'unités, y compris des manuscrits uniques ou des ouvrages dans un état matériel médiocre, consultés tous les jours par de plus en plus d'utilisateurs. Ces instruments de recherche analogiques sont aujourd'hui en grande partie informatisés (structurés en EAD) ; cependant un chantier de dématérialisation et d'encodage est encore en cours. Quant aux bases de données documentaires

archivistes-historiens[15]. **Ces inventaires manuscrits ou imprimés consistaient en bonne part en discours** composé de sections et de chapitres, de phrases et de mots parfois mis en forme, de notes et de références, associés à des éléments non rédigés (des données), dont le type de contenu et les règles d'organisation ont été fixés par ISAD(G)[16]. Le travail de mise en conformité avec ISAD(G) a permis d'organiser, de structurer plus précisément ce discours (en parties tels que l'historique de la conservation, les informations sur le classement, la présentation du contenu, ce pour chaque niveau de description dès lors que c'était pertinent). Alors que les premiers sites web culturels voyaient le jour, pour informatiser de telles informations, rien n'était plus naturel que le métalangage XML, qui est spécialement approprié pour structurer tout type de discours textuel, considéré comme une hiérarchie ordonnée d'objets de contenus, en utilisant des balises autodescriptives. **XML, et la grammaire EAD, ont été choisis parce qu'ils permettaient de conserver toute la substance de ces inventaires, de la pérenniser** (XML en tant que langage d'encodage indépendant des plates-formes et systèmes ; EAD en tant que grammaire XML publique et documentée), **de la diffuser sur internet, de la rendre exploitable et de l'échanger**. Il se trouve que XML, héritier d'un métalangage plus ancien, SGML (ISO 8879:1986), s'est trouvé rapidement accompagné d'une suite de technologies cohérentes et performantes[17]. Aujourd'hui, même si son audience s'est réduite, **XML reste la meilleure solution pour encoder tout type de discours structuré**. Il est notamment très utilisé par les chercheurs en sciences humaines et sociales pour produire et exploiter des éditions scientifiques, à grandes mailles ou beaucoup plus fines, de tout type de document textuel, du manuscrit médiéval au dictionnaire multilingue en passant par les séries de correspondances ou la transcription d'enregistrements sonores[18].

Cependant dans nos inventaires actuels, la place du discours s'est réduite, sous la pression de divers facteurs. Les grandes introductions, notamment, ne sont plus de mise. La relation entre archivistes et historiens s'est également distendue[19]. Mais cet héritage reste essentiel : nous conservons et utilisons encore parfois des inventaires analytiques ou états sommaires produits sous l'Ancien Régime ou au 19e siècle. Et surtout, **nos besoins d'information contextuelle nous conduisent encore et toujours à rédiger des textes parfois longs pour faire la présentation générale d'un fonds ou pour donner à comprendre, dans l'histoire d'un agent, où et quand il a existé, ce qu'il a fait et avec qui il était en relation**.

---

textuelles, une enquête avait alors été menée à bien, et un nombre important de bases de données locales, ou installées sur des serveurs du ministère et accessibles via des terminaux en mode texte, avaient été recensées.

[15] Voir à ce sujet l'article de Bruno Galland, La participation de l'archiviste à la recherche historique : un rôle à redéfinir ?, dans *La Gazette des Archives*, n° 204, 2006-4. Les archives en France, p. 211-232. DOI : https://doi.org/10.3406/gazar.2006.3837.

[16] La mise en œuvre d'ISAD(G) en France s'est appuyée sur le manuel publié en 1999 par Christine Nougaret et Bruno Galland*, Les instruments de recherche dans les archives* (Direction des Archives de France - La Documentation Française). Elle s'appuie également sur les éditions successives de l'*Abrégé d'archivistique* de l'AAF et sur quelques circulaires des Archives de France.

[17] Dont : les DTD, langages de schémas et Schematron pour modéliser les documents XML et en contrôler la qualité ; XPath pour parcourir les arbres XML ; XSLT pour transformer les documents XML ; XQuery pour les requêter. Ces technologies ont encore gagné en performance ces derniers temps (mention spéciale pour pour XSLT 2.0, et XSLT 3.0, qui, entre autres avancées, permet de transformer facilement du JSON en XML, ou du XML en JSON).

[18] Pour cela en général ces équipes et chercheurs utilisent le système descriptif TEI (https://tei-c.org/), dont la première version d'EAD a d'ailleurs été dérivée en 1998.

[19] Voir à ce sujet le bel article publié par Olivier Poncet, Archives et histoire : dépasser les tournants, dans *Annales. Histoire, Sciences Sociales*, 2019/3-4 (74e année), p. 711-743. DOI : https://doi.org/10.1017/ahss.2020.50.

Aucun graphe de connaissances ne peut remplacer pleinement de tels discours même plus courts et fragmentés ; plus exactement, si on voulait le faire (et cela se fera sans doute à moyen terme), il faudrait mobiliser à plein les méthodes et techniques du traitement automatique du langage et l'intelligence artificielle. Cependant, les graphes d'entités liées peuvent d'ores et déjà se nourrir de toutes les informations contenues dans ces inventaires[20] et, comme nous l'avons déjà dit, les lier entre eux, et même substituer au réseau des instruments de recherche et unités de description un réseau de ressources archivistiques (Record, Record Set, Record Part), chacun étant lié, au moins dans un premier temps, à un nombre n de descriptions plus ou moins anciennes issues de ces inventaires.

Au fond, franchir ce nouveau pas et choisir cette approche revient à **donner, ou rendre, à ces instruments de recherche leur véritable statut… celui d'archives (Record) au sens de RiC**, et à les décrire comme tels[21].

Les graphes de connaissances conformes à RiC, plus précisément à RiC-O puisqu'on commence à parler ici de réalisation concrète, pourraient alors venir, dans un premier temps au moins, se positionner "devant" les IR en XML, et les référencer ; ou encore (et ce n'est pas exclusif), de façon plus subtile, servir en arrière-plan pour étendre une recherche (en suivant les relations pour par exemple intégrer aux résultats les documents liés à celui qui a été trouvé, ou proposer de faire une recherche sur des producteurs ayant succédé à celui requêté, sur des lieux adjacents de celui initialement cherché, etc.).

Cette possibilité que les technologies de graphes nous offre est sans doute aussi **une opportunité pour recréer du lien entre l'archiviste et l'historien, contribuer par la co-construction, le partage ou le liage de données et de connaissances à rapprocher les deux communautés**[22]. À cet égard il faut noter que RiC propose un arsenal d'attributs et de relations permettant de consigner des informations de nature diplomatique (telles que le statut des documents, ou encore leur authenticité), de lier entre eux des documents génétiquement associés, de consigner la source utilisée pour établir certaines relations ou pour décrire des archives.

Quoi qu'il en soit, il convient d'avoir en tête que le format EAD, que beaucoup de services d'archives du monde utilisent, devrait évoluer dans les prochaines années, prenant notamment en compte, entre autres, RiC, pour aboutir en 2025 à une nouvelle version majeure, donc sans doute plus précise et permettant de lier plus facilement la description des documents avec celles des entités de contexte[23].

5. Est-ce que RiC reste pure théorie aujourd'hui, ou bien y a-t-il déjà des

[20] Rien n'empêche, par exemple, de stocker le texte d'une présentation de contenu dans l'attribut RiC-CM correspondant (Scope and Content), ou d'utiliser la datatype property correspondante dans RiC-O.

[21] RiC-CM et RiC-O 0.2. incluent aussi la relation 'describes o described' qui permet de lier une ressource archivistique à la ressource qui la décrit. Par ailleurs, RiC-O définit la catégorie de document (l'instance de la classe Documentary Form Type) Finding Aid. Il est d'ailleurs probable que dans RiC-O cette catégorie de document devienne une classe à part entière, sous-classe de Record.

[22] Comme O. Poncet, dans l'article déjà cité, l'appelle de ses vœux.

[23] C'est ce que prévoit le Technical Subcommittee on Encoded Archival Standards (TS-EAS) de la Society of American Archivists, qui est responsable de l'évolution d'EAD et d'EAC-CPF et qui communiquera plus précisément sur ce programme dès cette année.

réalisations concrètes, utilisant RiC-O et RDF ?

RiC-CM est un modèle conceptuel abstrait, donc oui en ce sens c'est de la théorie, mais élaborée par un groupe de praticiens experts du domaine[24], pour les besoins du métier et avec les communautés concernées de professionnels.

Rappelons qu'il s'agit aussi d'un brouillon (draft), dans lequel les réponses à l'appel à commentaires et les premières implémentations et leur suite permettront de dégager les composants les plus utiles. Nous avons aussi expliqué précédemment qu'il s'agit d'un modèle global. Cela signifie que dans le cadre d'une mise en œuvre locale, il n'est absolument pas obligatoire d'utiliser RiC dans sa totalité. Bien au contraire, on peut prévoir que la plupart des projets utiliseront un sous-ensemble de RiC. Par ailleurs, il est possible d'étendre RiC.

Donc RiC-CM, c'est par essence de la théorie. Mais il en existe d'ores et déjà des utilisations concrètes significatives. Et celles-ci se font souvent en utilisant RiC-O, la seule formalisation technique officielle existante de RiC-CM. Nous renvoyons pour commencer à la page sur les projets de mise en œuvre de RiC-O, que nous mettons à jour régulièrement[25].

Cette ontologie comme cela a été dit, sert à définir le vocabulaire et les règles pour produire, conformément à RiC-CM, des graphes de connaissances en RDF destinés à être publiés dans le web de données, ou web sémantique. **Le domaine ou le périmètre naturels de RiC-O, c'est le web, les interfaces et services web d'accès pour les utilisateurs finaux des métadonnées.**

Si l'on peut (s'il faut !) dès maintenant imaginer des trajectoires d'évolution de nos infrastructures souvent vieillissantes de stockage et de gestion des métadonnées archivistiques vers des solutions plus ouvertes orientées données, les graphes d'entités liées n'y seront probablement pas, le plus souvent, directement stockés et gérés via des bases RDF (triplestores). D'autres solutions sont souvent employées pour ce faire, de la base de données relationnelle aux bases NoSQL, qui sont connues pour leur robustesse et leur performance. **RiC-O ne constitue pas du tout une injonction à utiliser RDF pour le cœur des SIA**. Répétons-le, RiC-O est simplement et avant tout le modèle qui manquait aux professionnels des archives pour restructurer puis pousser dans le web de données, donc rendre accessibles aux humains et exposer aux machines dans différents formats (XML, Turtle, JSON-LD), des jeux de données de qualité, les lier entre eux, ou avec d'autres jeux de données.

Une telle démarche n'est d'ailleurs pas du tout exclusive d'autres démarches. Ce qui peut être exposé sous forme de simples fichiers tabulés ou CSV doit continuer à l'être. Disposer, dans son système d'information local, d'un entrepôt OAI-PMH exposant des métadonnées archivistiques en XML/EAD ou en Dublin Core qualifié issu du XML/EAD, par ex., permet d'ores et déjà à des agrégateurs comme FranceArchives de moissonner l'ensemble de ces métadonnées (ou des sous-ensembles prédéfinis).

---

[24] La grande majorité d'entre eux sont des praticiens exerçant des missions à haut niveau de responsabilité (directeurs de services d'archives ou de gros projets numériques comme SNAC, gérant de SSII spécialisée) ; quelques-uns sont enseignants chercheurs.

[25] https://ica-egad.github.io/RiC-O/projects-and-tools.html. Il faut noter que l'équipe du portail FranceArchives a choisi, pour sa future version, d'utiliser RiC-O.

Rien n'empêche par ailleurs de publier directement ses fichiers EAD ou EAC(CPF), ou son registre des entrées, en open data. On publiera alors des fichiers conformes à plusieurs modèles différents.

Pour les institutions et projets qui souhaiteraient utiliser RiC-O, on voit plutôt se dessiner des évolutions consistant, non pas à remplacer ce qui existe par des entrepôts RDF, mais à ajouter aux SIA actuels les briques nécessaires pour générer et exposer de tels jeux de métadonnées.

On peut s'inspirer de l'exemple de la BnF. La BnF développe depuis dix ans data.bnf.fr . **Cette application est construite avec le logiciel CubicWeb (cf. https://fr.wikipedia.org/wiki/CubicWeb), qui emploie entre autres une base de données relationnelle pour stocker les données**. Les données de data.bnf.fr sont le résultat du regroupement et du retraitement de métadonnées de structures et formats divers produites par la BnF, pour les rendre conformes à un modèle unique fondé sur IFLA LRM. Data.bnf.fr génère dynamiquement, en réponse aux requêtes des utilisateurs finaux, des vues de ces données qu'il embarque dans les pages HTML retournées au navigateur web des utilisateurs. Les triplets RDF correspondant à chaque page web peuvent être téléchargés depuis la page dans les différents formats de sérialisation de RDF. En outre la source HTML de ces pages web de data.bnf.fr embarque des données RDFa conformes aux modèles génériques Schema.org et Opengraph Protocol, ce qui permet aux moteurs de recherche du web d'indexer finement ces pages web. Enfin data.bnf.fr est dotée d'un triplestore et d'un SPARQL endpoint, par lequel elle rend la version RDF de l'ensemble de ces données interrogeable directement par les utilisateurs finaux, ou à distance pour les machines (SPARQL étant en même temps un langage de requête pour les données RDF et un protocole de recherche)[26]. La BnF propose ainsi une API d'accès à une version RDF de ses données, parmi d'autres APIs[27]. Signalons au passage que la BnF a inscrit dans la feuille de route data.bnf.fr pour 2020-2024 l'intégration de RiC-O pour un meilleur traitement des données issues de la Base Archives et Manuscrits (BAM)[28].

De plus, la BnF et l'ABES élaborent en ce moment le Fichier National d'Entités (FNE), une plateforme de production mutualisée de descriptions d'entités structurées conformément à IFLA-LRM[29].Pour le cœur de cette plateforme, le choix logiciel s'est porté sur la plateforme Wikibase, utilisée désormais par un nombre significatif d'institutions et de projets, dont l'architecture inclut un triplestore RDF (Blazegraph) adossé à une base de données relationnelle (cf. https://addshore.com/2018/12/wikidata-architecture-overview-diagrams/).

On voit bien, dans ces exemples concernant le web de données, que **les bases de données (triplestores) RDF s'inscrivent dans des écosystèmes vastes, dans lequel elles jouent un rôle précis, et qu'elles n'y servent pas de moyen de stockage et d'administration des métadonnées**.

Quels que soient les choix d'architecture techniques à faire et à implémenter dans les systèmes d'information locaux si l'on souhaite « passer à RiC » et publier les résultats dans le web de données, il s'agit d'un important travail (comme pour toute implémentation industrielle), qui devrait à notre sens aux AN commencer par les référentiels, en tant que clés de liage internes et externes des

[26] Nous résumons ici ce qui est expliqué aux pages suivantes : https://data.bnf.fr/fr/about et https://data.bnf.fr/semanticweb.

[27] Voir https://api.bnf.fr/fr/, et au sujet du service SPARQL d'accès, la page https://api.bnf.fr/fr/sparql-endpoint-de-databnffr. Par ailleurs bien sûr les jeux de données ouvertes RDF de la BnF sont directement téléchargeables (https://api.bnf.fr/fr/dumps-de-databnffr).

[28] Voir https://data.bnf.fr/roadmap.

[29] https://www.transition-bibliographique.fr/fne/fichier-national-entites/.

métadonnées, et compte tenu des problèmes de modèle, de gestion et de positionnement qu'ils cumulent.

Mais en attendant, rien n'empêche, même si on ne dispose pas d'application comme data.bnf.fr, pleinement intégrée au SI institutionnel, de commencer à produire des métadonnées conformes à RiC, puisqu'on dispose de cette ontologie RiC-O, projet de standard lui-même fondé sur un modèle conceptuel international et des langages informatiques publics dotés de spécifications.

Aux AN nous avons donc produit, et mettons à jour régulièrement, des jeux de données RDF conformes à RiC-O à partir :

- **de l'ensemble des instruments de recherche en EAD** (soit actuellement environ 29000 fichiers) ;

- **de l'ensemble des notices d'autorité de producteurs en EAC-CPF** (soit actuellement environ 15210 fichiers) ;

- **de l'ensemble des vocabulaires servant actuellement à l'indexation des inventaires et notices de producteurs** (soit une vingtaine de vocabulaires).

Cela constitue en soi un pas en avant très concret et important, par rapport à l'étape précédente pendant laquelle nous avions élaboré, en collaboration avec la BnF, le SIAF et la société Logilab, la preuve de concept qualitative PIAAF (https://piaaf.demo.logilab.fr, publiée en février 2018). En effet nous sommes ainsi passés à une échelle industrielle et systématique.

Nous faisons cela automatiquement, en utilisant, d'une part, pour les deux premiers ensembles, le logiciel open source RiC-O Converter[30] développé par la société Sparna et publié en avril 2020 ; d'autre part, une série de scripts XSLT que nous avons écrits et maintenons (la structure des référentiels d'indexation des AN étant appelée à évoluer fortement, nous avons dans l'immédiat préféré cette solution). Nous souhaitons maintenant commencer à travailler à la conversion en RDF/RiC-O des fichiers SEDA du SAE des AN, ce qui nous conduira très certainement à envisager d'associer des données RIC-O et des données conformes à l'ontologie PREMIS.

Dans le cadre du projet RiC-O Converter, **nous avons d'ailleurs produit et publié, dans la documentation du logiciel, des mappings (tableaux de correspondance) entre éléments EAD et composants de RiC-O, éléments EAC-CPF et composants (classes et propriétés) de RiC-O[31]**. Même s'ils ne prennent pas en compte la totalité des éléments ou attributs des formats source mais seulement ceux utilisés par les AN**, ces mappings**, les seuls qui existent à notre connaissance pour l'instant, **peuvent servir de base de travail à n'importe quel autre projet et institution souhaitant convertir en RDF/RiC-O des fichiers EAD ou EAC-CPF**, qu'elle utilise RiC-O Converter ou non. Nous allons par ailleurs réaliser avec Sparna une version 2.0 de RiC-O Converter et de sa documentation cette année, pour les rendre conformes à RiC-O 0.2 publiée en février 2021.

---

[30] https://github.com/ArchivesNationalesFR/rico-converter. Au sujet de RiC-O Converter, voir aussi la présentation qui en a été faite par Thomas Francart et Pauline Charbonnier le 28 janvier 2020 (https://labarchiv.hypotheses.org/files/2020/02/20200128_4_RiCOConverter.pdf) et les enregistrements vidéo du webinaire organisé en juin 2020 avec l'aide de l'AAF (https://www.dailymotion.com/playlist/x6x1d0).

[31] Ce sont les deux fichiers Excel disponibles dans le dossier https://github.com/ArchivesNationalesFR/rico-converter/tree/master/ricoconverter/ricoconverter-doc/src/main/resources.

Voilà qui prouve au moins de façon très claire qu'on peut bel et bien produire du RDF conforme à RiC-O à partir d'un corpus de fichiers EAD (d'instruments de recherche archivistiques classiques) et de notices EAC-CPF tels que produits par un SI institutionnel.

Si c'est **un pas en avant important, c'est aussi parce que cela oblige à s'interroger sur ce dont on dispose véritablement, sur la cible qu'on souhaite atteindre, sur la distance à parcourir pour atteindre la cible et sur les méthodes à mettre en œuvre pour atteindre cette cible**. Toutes réflexions qui sont très importantes pour la suite, et doivent nécessairement être menées avant de toucher au SI. Une série de questions doit donc être traitée :

- En ce qui concerne l'existant : de quoi est-il constitué exactement ? Si, pour les notices de producteurs et les vocabulaires les AN ont une vision assez claire de ce qu'elles ont produit, pour les 29000 instruments de recherche, c'est plus difficile à dire, étant donné le caractère plastique d'EAD, la grande diversité de nature des inventaires et des documents décrits, et la quantité de métadonnées. À l'occasion du projet RiC-O Converter, en préparant les tests unitaires, nous avons ainsi, par exemple, fouillé assez systématiquement les fichiers EAD et nous percevons mieux les différents choix d'encodage qui ont été faits pour un élément EAD ou un autre, le caractère parfois très faiblement structuré ou plus rarement les écarts constatés entre certains usages et les règles énoncées par la documentation EAD.

Le diagnostic n'est pas achevé aux AN, et le projet RiC-O Converter n'est évidemment pas le seul à pouvoir apporter des éclairages sur ce sujet. **Mais, en tenant compte de l'objectif de construction d'un graphe d'entités liées, il est déjà certain qu'une marge de progression qualitative existe**, qui demande, pour être réduite, d'élaborer un plan collectif d'action, associé à des modifications techniques et fonctionnelles du SI : une plus grande attention à la rédaction des IR, des extensions au profil d'application EAD des AN trop restrictif, parfois des contraintes plus fortes, un effort d'indexation (humaine ou semi-automatique), des évolutions importantes du contenu et de la structure des référentiels.

- En ce qui concerne la cible : à partir de l'existant, quelles catégories d'entités (instances de classes RiC-O) veut-on, et peut-on dans l'immédiat, envisager d'isoler et de décrire ? Avec quels attributs/propriétés ? Quelles relations distinguer entre les entités ? **Cela pose la question de la granularité des données à produire, et cela renvoie à la notion de ce qui est utile, ou suffisant (« good enough ») pour répondre aux besoins utilisateurs mentionnés précédemment**. Une autre question, tout aussi importante, est celle de l'assignation d'identifiants uniques et pérennes aux entités qu'on va décrire en RDF/RiC-O ; nous ne ferons que la mentionner ici.

Il faut donc faire des choix, l'idée étant de tirer le meilleur parti de l'état actuel des métadonnées et de prendre en compte toutes les modalités d'utilisation des éléments ou attributs XML, donc divers scénarios de conversion, de la plus simple à la plus élaborée possible.

Dans l'entrepôt GitHub de RiC-O, on trouvera quelques exemples de fichiers EAD et de notices EAC-CPF des AN, avec les fichiers RDF/RiC-O v0.2 issus de la conversion[32]. On trouvera aussi d'autres exemples préparés à partir de métadonnées produites hors de France, tels qu'un ensemble de fichiers EAD et EAC-CPF exportés depuis une application web fondée sur le logiciel ICA ATOM, et un exemple

[32] Voir https://github.com/ICA-EGAD/RiC-O/tree/master/examples/examples_v0-2/NationalArchivesOfFrance.

très simple, mais éclairant, de triplets RDF utilisant RiC-O et l'ontologie PREMIS pour décrire un document nativement numérique.

6. Mais finalement à quoi RDF/RiC-O peut-il très concrètement être utile pour un service d'archives ?

Il est bien sûr plus que légitime de se poser cette question.

Nous avons apporté des éléments de réponse lorsque nous avons évoqué l'intérêt du passage à RiC. **Pour être un peu plus concret maintenant, revenons au cas des AN, et parlons d'utilisation de métadonnées RDF/RiC-O et des technologies RDF**. Nous évoquerons d'autres cas que celui des AN plus tard dans ce carnet de recherche.

Même si le SI institutionnel des AN n'intègre pas encore de brique sémantique, **les jeux de données RDF de taille conséquente dont nous disposons servent déjà, de multiples manières**. Nous prendrons seulement ici deux exemples.

Tout d'abord parlons d'un projet de recherche piloté par l'IGN, financé par l'ANR et qui arrive à son terme fin 2021, ALEGORIA[33]. ALEGORIA concerne des fonds et collections de photographies aériennes et autres documents graphiques ou iconographiques relatifs au territoire français, conservés par différentes institutions (principalement l'IGN, le Musée Nicéphore Niepce, les AN), décrits selon des modèles et des formats divers. L'objectif du projet était d'en réunir les reproductions numériques et les métadonnées au sein de la même plateforme web, et d'utiliser conjointement les technologies RDF et l'intelligence artificielle (moteur de recherche multimodal combinant recherche par métadonnées et recherche sur le contenu des images) pour chercher, rapprocher des documents, déduire des regroupements obtenus des données manquantes (date d'un cliché, lieu représenté…), charger les documents retenus par l'utilisateur au sein d'une plateforme 3D immersive. Même s'il s'agit d'une preuve de concept, **la plateforme web qui sera publiée cet automne intègrera environ 40000 photographies et documents, tous décrits, quel que soit leur lieu de conservation, en utilisant RiC-O et RDF, ainsi que des référentiels liés entre eux** (en particulier les référentiels de lieux de l'IGN et ceux des AN) ; ce graphe de données sera téléchargeable et interrogeable à distance via SPARQL. **Pour les AN, parmi les résultats directement et immédiatement utilisables du projet, on peut mentionner : des instruments de recherche en EAD au contenu affiné et indexé, des référentiels de lieux fortement enrichis, et on le souhaite, des métadonnées complémentaires à celles existant aux AN avant le projet**.

[33] Voir le site web du projet : https://www.alegoria-project.fr/projet.

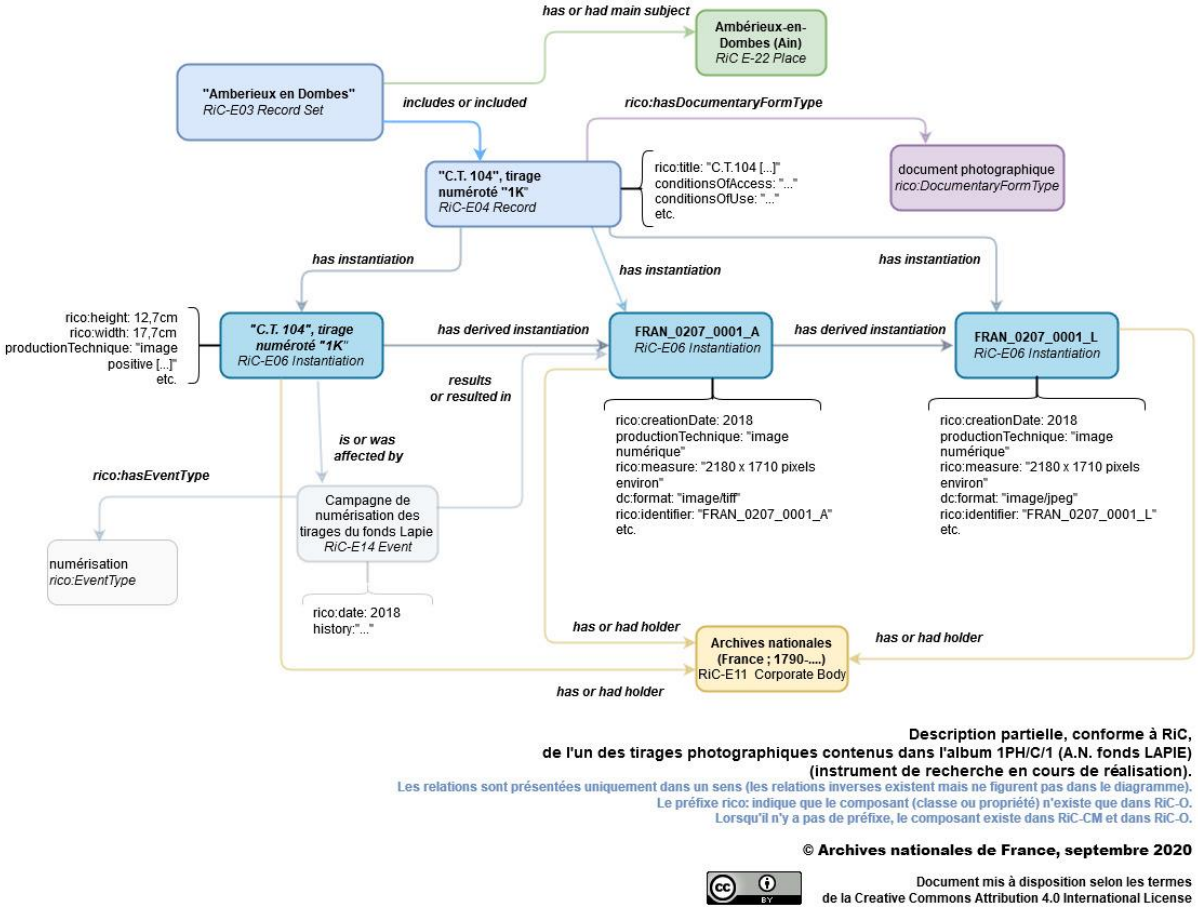


Figure 3 : exemple de mise en œuvre de RiC-O 0.2 dans le cadre du projet ALEGORIA

Ce projet est certes très spécifique, mais aussi assez représentatif de plusieurs autres projets de recherche moins avancés, qui impliquent les AN comme partenaire ou qui sont menés directement par l'institution, et dans lesquels des graphes de données liées utilisant RiC-O vont être produits. Dans tous les projets un travail préalable de modélisation et de traitement de métadonnées est nécessaire, qui ne peut pas être réalisé via les outils mentionnés plus haut (RiC-O Converter notamment), en raison de la précision demandée. Dans tous les projets on espère des retombées significatives en ce qui concerne la qualité des métadonnées initialement mobilisées, la consolidation des référentiels locaux, peut-être des « découvertes » scientifiques pour les chercheurs impliqués et pour les archivistes, et la publication de données liées (Linked Open Data), sans parler de la mise à l'épreuve de RiC.

Le deuxième exemple que nous prendrons concerne plus globalement l'institution. Il s'agit de la construction des référentiels, sur laquelle nous nous efforçons d'avancer actuellement, avec les moyens du bord, considérant qu'il s'agit vraiment d'un enjeu essentiel de la transition amorcée. **Sans les technologies RDF et notamment SPARQL, nous n'aurions pas pu intégrer, dans la description relative aux circonscriptions administratives contemporaines, les données historiques contenues dans les jeux de données de l'INSEE, ni les données de géolocalisation extrêmement précises contenues dans les jeux de données de l'IGN. Nous n'aurions pas pu non plus commencer à aligner certains référentiels, comme celui des personnes physiques ou des personnes morales, sur Wikidata. Sans RDF, RiC-O et SKOS, nous aurions sans doute eu du mal à concevoir rapidement un modèle de données riche pour ces référentiels, correspondant à la nature des entités décrites**.

Au bout du compte, comme ces référentiels étaient nécessaires à plusieurs projets (dont un projet de reconnaissance semi-automatique d'entités nommées) et comme nous avions besoin de les versionner, nous avons choisi de les publier sur GitHub, dans leur format source pour certains, dans le format RDF/RiC-O pour tous, dans leur état actuel et malgré leurs défauts[34]. Nous communiquerons plus précisément à ce sujet lorsque nous aurons tout chargé dans le dépôt GitHub et produit une documentation. Nous pensons bien sûr que ces référentiels pourraient servir à d'autres que leurs premiers destinataires, même en l'état. Dans le cas où des réutilisations seraient envisagées, les IRIs fournis ne sont pas déréférençables actuellement.

De manière plus générale, on sait qu'utiliser SPARQL pour requêter des graphes de données RDF permet d'interroger et de parcourir les relations liant les entités décrites, pas seulement leur description**. Ce qui nous conduit à affirmer que des questions qui ne peuvent pas être posées dans un SI archivistique classique peuvent alors être posées**.

Ainsi, dans le cas des AN, nous savons depuis longtemps que les fichiers EAD et EAC-CPF forment, par le jeu des relations EAC-CPF essentiellement (qui relient entre elles les notices d'autorité ou les notices d'autorité et les instruments de recherche) un réseau socio-documentaire très dense[35]. Mais ce réseau n'est que très peu visible dans l'actuelle SIV, et il ne peut pas être interrogé. Ainsi, il n'est pas possible de demander par exemple, la liste des services d'administration centrale qui ont succédé au premier bureau des monuments historiques du ministère de l'intérieur jusqu'à nos jours, en triant le tout dans l'ordre chronologique de création de ces organismes (et optionnellement, les archives que ces services ont produites). Alors qu'avec SPARQL, une fois les données RDF des AN chargées dans un triplestore, on peut poser cette question, qui a bien évidemment un sens, notamment pour l'archiviste, pour le chercheur en histoire des institutions ou de l'art.

La question peut se poser ainsi :

[34] https://github.com/ArchivesNationalesFR/Referentiels.
[35] Voir à ce sujet Article Gazette des archives

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rico: <https://www.ica.org/standards/RiC/ontology#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
SELECT ?entite ?nom ?dateDebut ?dateFin ?archives WHERE
{
      ?bureauMH rdf:type rico:CorporateBody; rdfs:label ?label .
    FILTER contains(?label, 'Monuments historiques (1839'@fr) .
    ?bureauMH (rico:precedesInTime)+ ?entite .
    ?entite rdfs:label ?nom; rico:beginningDate ?dateDebut .
    OPTIONAL {
        ?entite  rico:endDate ?dateFin  }
    OPTIONAL {
    ?entite rico:isProvenanceOf ?archives  }
}
ORDER BY ?dateDebut
```

Figure 4 : une requête SPARQL parcourant un graphe de données RiC-O

Elle retourne, dans l'actuel graphe des AN, la liste de réponses ci-après :

	entite	nom	dateDebut	dateFin
1	http://data.archives-nationales.culture.gouv.fr/agent/051121	"France. Ministère de l'Intérieur. Direction des Beaux-Arts. Bureau des Beaux-Arts, des Musées et des Monuments historiques (1848-1852)"@fr	"1848-01-01"^^xsd:date	"1852-12-31"^^xsd:date
2	http://data.archives-nationales.culture.gouv.fr/agent/051122	"France. Ministère d'État. Division des Beaux-Arts. Service des Monuments historiques (1853-1858)"@fr	"1853-01-01"^^xsd:date	"1858-12-31"^^xsd:date
3	http://data.archives-nationales.culture.gouv.fr/agent/051123	"France. Ministère d'État. Bureau des Monuments historiques (1858-1863)"@fr	"1858-01-01"^^xsd:date	"1863-12-31"^^xsd:date
4	http://data.archives-nationales.culture.gouv.fr/agent/051124	"France. Ministère de la maison de l'Empereur. Surintendance des Beaux-Arts. Bureau des Monuments historiques (1863-1870)"@fr	"1863-01-01"^^xsd:date	"1870-12-31"^^xsd:date
5	http://data.archives-nationales.culture.gouv.fr/agent/051125	"France. Ministère de l'Instruction publique. Direction des Beaux-Arts. Bureau des Monuments historiques (1870-1907)"@fr	"1870-01-01"^^xsd:date	"1907-12-31"^^xsd:date
6	http://data.archives-nationales.culture.gouv.fr/agent/051126	"France. Ministère de l'Éducation nationale. Direction des Beaux-Arts. Bureau des Monuments historiques (1907-1939)"@fr	"1907-01-01"^^xsd:date	"1939-12-31"^^xsd:date
7	http://data.archives-nationales.culture.gouv.fr/agent/051149	"France. Ministère de l'Éducation nationale. Secrétariat général des Beaux-Arts. Direction des services d'architecture. Bureau des Monuments historiques et Sites (1940-1944)"@fr	"1940-01-01"^^xsd:date	"1944-12-31"^^xsd:date
8	http://data.archives-nationales.culture.gouv.fr/agent/051150	"France. Ministère de l'Éducation nationale. Direction de l'architecture. Sous-direction des Monuments Historiques et des Sites (18 août 1945-8 janvier 1959)"@fr	"1945-08-18"^^xsd:date	"1959-01-08"^^xsd:date
9	http://data.archives-nationales.culture.gouv.fr/agent/051151	"France. Ministère de la Culture et de la Communication. Direction de l'Architecture et du Patrimoine. Sous-direction des Monuments Historiques et des Espaces protégés (1959-2009)"@fr	"1959-01-08"^^xsd:date	"2009-12-31"^^xsd:date
10	http://data.archives-nationales.culture.gouv.fr/agent/051153	"France. Ministère de la Culture et de la Communication. Direction générale des patrimoines. Service du patrimoine. Sous-direction des Monuments historiques et des Espaces protégés (2009-....)"@fr	"2009-01-01"^^xsd:date	

Figure 5 : la liste de résultats répondant à la requête SPARQL présentée figure 4

Comme il s'agit d'explorer un graphe de connaissances, cela ouvre de nouvelles perspectives et des découvertes pourraient donc être faites par les utilisateurs finaux.

La présentation du logiciel RiC-O Converter faite en janvier 2020 donne d'autres exemples de requêtes SPARQL. Elle laisse aussi entrevoir une interface intuitive, souple et réactive de construction de

requêtes, construite à l'aide du composant Javascript open source Sparnatural[36]. Implémenter un tel outil permet de lever deux verrous bien connus des organismes ou projets utilisant RDF, qui constituent des freins importants à son adoption :

- l'utilisateur de l'interface n'a pas besoin de connaître SPARQL pour construire sa requête ;

- il n'a pas besoin de connaître l'ontologie de référence non plus.

Nous allons travailler cette année avec la BnF et le ministère de la Culture (qui nous apporte notamment un soutien financier indispensable), dans le cadre d'un marché public, à améliorer Sparnatural. À la fin du projet les trois partenaires (et plus largement toute institution ou équipe puisqu'il s'agira d'un produit open source) disposeront d'un outil facile à configurer selon l'ontologie de référence et à intégrer à une application web ; de plus deux démonstrateurs auront été élaborés en associant un panel d'utilisateurs finaux. Le démonstrateur des AN devrait intégrer une quantité significative des données RDF de l'institution.

Nous aurons ainsi franchi une autre étape des travaux visant au final à mettre en place une application sémantique pour l'institution, ce qui impliquera de revenir vers le SI actuel et de programmer diverses actions. Nous n'oublions pas la problématique de la visualisation des données RDF, à laquelle nous avions commencé à réfléchir dans le cadre du projet PIAAF.

Il ne s'agit que de quelques exemples d'utilisation de RDF, qui ne préjugent pas de l'avenir ni d'autres possibilités.

7. En conclusion (de cet article)

Aucune technologie n'est parfaite pour tout faire. De nombreuses années d'expérience dans ce domaine, au carrefour de l'archivistique et des sciences de l'information, nous ont appris qu'il faut savoir tirer le meilleur parti de chacune. Par ailleurs, bien sûr, une norme n'est qu'un point de départ. Et bien sûr, la rigueur, la lucidité et le pragmatisme sont indispensables lorsqu'on veut proposer et mettre en œuvre une stratégie pour faire évoluer à l'échelle d'une institution ou d'un projet des jeux de métadonnées massifs, dont la production demande des efforts et des qualifications de haut niveau à des collectifs de personnes, ou pour transformer les architectures techniques actuelles. Un autre élément essentiel de ces processus c'est le temps ; il en faut. Enfin, aucun projet de ce type ne réussit sans l'adhésion et les apports de tous, sans la volonté de l'institution et sans une organisation durable.

Ceci étant dit, nous sommes optimistes, comme d'autres personnes, sur l'avenir des graphes de connaissances RDF. Il ne s'agit pas uniquement d'entrepôts de stockage, mais d'un ensemble de technologies, incluant aujourd'hui un éventail de standards du W3C dotées de spécifications publiques (RDF, RDFS, OWL, SPARQL), divers formats de sérialisation, des briques logicielles diverses avec une offre variée. Des nouveautés importantes sont venus compléter cet ensemble ces dernières années :

[36] Voir le site web du projet : http://sparnatural.eu/. On trouve aussi sur la page https://api.bnf.fr/fr/node/158 un lien vers une petite vidéo de démonstration de l'utilisation de Sparnatural pour interroger les données RDF de data.bnf.fr.

SHACL[37], un langage pour contrôler la qualité des graphes RDF ; plus récemment, RDF star[38], dont les spécifications ne sont pas complètes mais qui pourrait sans doute apporter des solutions intéressantes pour décrire des triplets. Au moyen de diverses technologies incluant ces technologies RDF, de grands réservoirs (notamment des référentiels comme le FNE) sont en cours de construction et de publication, qui sont ou seront sans nul doute incontournables. Nous croyons aussi à la créativité des ingénieurs connaissant ces technologies pour développer des interfaces de recherche et de consultation, mais aussi des services web ou des APIs.

Plus fondamentalement, s'il y a un point qui fait l'objet globalement d'un consensus, c'est que ces technologies sont une solution performante pour mettre en place des dispositifs d'interopérabilité au niveau fin de la donnée, consolider des données locales et interconnecter des données distribuées. Or nous croyons que **l'interopérabilité entre systèmes et le partage des données vont se développer, sans doute lentement, entre les institutions culturelles, et entre les institutions culturelles et les communautés d'utilisateurs.**

Et à notre avis, **c'est pour les archives un enjeu crucial que l'interopérabilité**. Voici sans prétendre à l'originalité ou à l'exhaustivité une liste de quelques arguments :

- **Chacun de nos services d'archives se situe dans un écosystème (où l'on trouve par exemple des services versants, et le portail FranceArchives) dans lequel l'échange de données est une nécessité** ; ainsi par exemple, on peut souhaiter que les missions d'archives dans les ministères accèdent enfin, depuis les applications qu'elles utilisent pour décrire les archives qu'ils conservent, aux référentiels des AN afin de pouvoir indexer les métadonnées descriptives qu'elles fournissent aux AN.

- **Nos référentiels (nos métadonnées de référence) sont incomplets et nous avons besoin de les lier à d'autres pour les enrichir ; inversement ce qu'ils contiennent est souvent d'un grand intérêt pour d'autres communautés, voire quasiment absent de toute autre base de connaissances**. C'est le cas notamment assez souvent pour nos notices de producteurs ; mais nous pourrions en dire autant pour d'autres entités telles que les types de documents.

- **Rappelons une évidence : les gisements documentaires que nous conservons sont également incomplets, dispersés, complémentaires et parfois redondants**, que ce soit du fait des aléas de l'histoire humaine (démembrements, transferts, destructions…) ou parce que les processus métier conduisent à des interventions simultanées ou successives de plusieurs personnes et collectivités selon leurs fonctions, donc à la production de différents dossiers 'locaux' aux contenus liés, et/ou à la transmission de documents et données entre agents. La question d'une transformation des métadonnées archivistiques est liée à celle de la macro-évaluation des archives ; alors lier à terme ces dossiers ou archives dispersées serait extrêmement utile, y compris pour les archivistes ;

- **Les utilisateurs des archives que nous conservons (notamment les chercheurs, associations, communautés citoyennes…) peuvent nous aider en produisant des données liées aux nôtres ; par ailleurs ils réclament de la donnée ouverte, réutilisable, clairement structurée et riche**.

Ainsi, soyons positifs et ayons un peu d'audace. **Essayons**, en nous fondant sur une expertise métier solidement établie et sur notre mission de service public, **de voir un peu loin et d'engager nos institutions dans cette transformation, sans jamais perdre de vue qu'il s'agit avant tout de stratégie**

[37] Shapes Constraint Language ; voir https://www.w3.org/TR/shacl/.
[38] Voir https://w3c.github.io/rdf-star/.

scientifique, et que les choix technologiques doivent toujours avoir du sens, et sans oublier que nos systèmes d'information sont de toute façon sujets à obsolescence et que la qualité de nos métadonnées est un point essentiel pour les générations actuelles et à venir. Il est très souhaitable que le réseau des archivistes français s'empare de ce sujet, réfléchisse à ce qu'il peut faire de RiC, et se fasse entendre. Au final, il ne peut en sortir que du mieux.